Research Statement

Bingwen Lu, Ph.D.

My research interest is to develop computational methods for mass spectrometry proteomics data analysis and to apply such methods to elucidate cell signaling pathways in human cancer cells. I have developed a series of software tools for mass spectrometry and proteomics data analysis. These tools improve de novo peptide sequencing[1,2] and mass spectrometry database search[3], increase the sensitivity of protein identification[4], evaluate the validity of phosphopeptide identifications[5], and screen phosphopeptide spectra prior to database search[6]. My current and future research is focused on the following areas: (1) developments of new tools for proteomics data analysis; (2) using existing as well as new tools in a systems biology approach to elucidate cell signaling pathways in human cancer cells.

(1) Computational Tools development

I am interested in developing tools in these areas: (a) phosphoproteomics data analysis and applications; (b) high mass accuracy proteomic mass spectrometry data analysis and applications; (c) differential proteomics study.

(a) Phosphoproteomics data analysis: Protein phosphorylation is an important event in the cells. Precise understanding of protein phosphorylation and localization of phosphorylation sites are of great interest to many biological and medical researchers. Until recently, identification of phosphopeptides and phosphoproteins by mass spectrometry, and the localization of phosphorylation sites, have been largely addressed by laborious manual validation. Automatic analysis, interpretation, and validation remain major logistical challenges when attempting to identify a large number of protein phosphorylation sites by tandem mass spectrometry experiments. To tackle the challenges, I designed a support vector machine-based algorithm (DeBunker) for automatic validation of phosphopeptides identified from tandem mass spectra[5] and designed an algorithm for automatic screening of phosphopeptide spectra prior to database search that reduced 60%-90% of computational time spent on database search[6]. These tools have been successfully applied in large scale phosphoproteomics data analysis[7-11]. Nevertheless, there are still many problems remained to be solved for phosphoproteomics data analysis, such as quantification of differentially phosphorylated peptides between control and treated samples and precise localization of phosphorylation sites and other post-translation modifications (such as acetylation, methylation, and ubiquitination). I will further develop computational methods and tools to tackle such problems further apply these tools to understand cell signaling pathways.

(b) High mass accuracy proteomic mass spectrometry data analysis: Traditional ion trap mass spectrometers have high sequencing speed and high sensitivity of ion detection. However, they tend to have relatively low mass accuracy measurement. Fourier transform (FT) ion cyclotron resonance (ICR) is a type of high resolving power mass spectrometer that has existed for several decades. However, FT-ICR mass spectrometers were not routinely used in proteomics study due to their high costs and maintenance demands. The linear ion trap (LTQ)-Orbitrap is a hybrid mass spectrometer that combines the efficiency and sensitivity of the linear ion trap with the high mass accuracy and high resolution of the Orbitrap mass analyzer. The LTQ-Orbitrap has been shown to routinely achieve sub-5-ppm mass accuracy at a dynamic range of more than 5000 and thus is helpful in various aspects of proteomics research. I have shown that high accurate mass measurement can be employed to improve protein identification sensitivity by combining both MS and MS/MS data[4]. High mass accuracy measurement of tandem mass spectra also assisted obtaining high confidence amino acid sequence de novo[12]. The high accurate mass measurement can also be used for the validation of cross-linking peptides[14]. Clustering high mass accuracy tandem mass spectrometry data can aid in the identification of novel peptides and protein post-translational modifications. Besides the above applications, high mass accuracy mass spectrometry data can also be utilized for the analysis of nuance protein post-translational modifications, alternative splicing, and amino acid polymorphisms, all of which are believed to play important roles in the cause of human cancers. I will further develop computational methods and tools for solving such problems.

(c) Differential proteomics study: The comprehensive analysis of protein expression in complex biological systems has demanded the development of new technologies to study proteomes of cells or organisms. Differential proteomics, the comparison of distinct proteomes (normal versus drug-treated cells, normal versus diseased tissues, etc.), is of paramount importance in elucidating of molecular and cellular mechanisms, including cancer mechanisms. Liquid chromatography coupled to tandem mass spectrometry has emerged as the main technology for large-scale in-depth differential proteomics analysis. Due to the large amount of mass spectra data generated by this method, biological inference poses a big challenge. We have identified key kinases that regulate HeLa cell adhesion and mobility[9, 11] using stable isotope labeling method. I have also used label-free quantitation methods to find proteins differentially expressed in the saliva of female human subjects selected from different age groups[13]. Currently, we are also designing methods for combining proteomics data with micro-array data to study different states of the cells. I will continue to develop tools for differential proteomics study and biological discoveries.

(2) Systems biology approaches to study cancer cell signaling pathways

I am interested in applying systems biology approaches in elucidating cancer cell signaling pathways. Specifically, I am using the human cell lines (HeLa, PC3, and MB231) as my model systems and I am interested in studying cell signaling pathways during cancer metastasis.

The most common cancers in human include breast cancer, prostate cancer, lung cancer, colon cancer, and ovarian cancer, and their metastasis is the leading cause of mortality in cancer patients, causing 90% of deaths from solid tumors (Gupta et al., 2006). During the process of metastasis, tumor cells leave the primary site, travel via blood and/or lymphatic circulatory systems, attach to the substratum of extracellular matrix (ECM) at a distant site, and establish a secondary tumor, accompanied by angiogenesis of the newly formed neoplasm (Gupta et al., 2006). Two of the critical steps of metastasis are adhesion and migration of the primary tumor cell on the ECM at the distant site.

During metastasis, cell attachment to, spreading on, and movement along the ECM are integrin dependent. Integrins are a large family of transmembrane receptors consisting of two heterodimeric, noncovalently-linked alpha and beta chains. Eighteen alpha chains and eight beta chains have been identified and different pairing of alpha and beta chains can yield more than 20 receptors that bind to different ECM proteins including collagen, fibronectin and laminin. Integrins play indispensable roles in delivering extracellular signals across the cell membrane into the cell interior through various kinases and phosphatases which modulate the phosphorylation status of their targets (Miranti et al., 2002). Understanding the mechanism by which integrins modulate these cellular activities is of significant biological importance.

In a preliminary study, we employed a systems biology approach using SILAC (Stable-Isotope Labeling by Amino acids in Cell culture) and MudPIT (Multi-dimensional Protein Identification Technology) to characterize the integrin-modulated phosphoproteomic network[9]. Using such an approach, we identified 357 proteins with significant phosphorylation changes during cell adhesion to, and spreading on, type I collagen. Among these proteins, 33 key signaling mediators with kinases or phosphatase activity were subjected to siRNA-based functional screening resulting in three integrin-regulated kinases, DBF4, PAK2, and GRK6.

For my future study, I will continue to use a combination of SILAC and MudPIT, together with proteomics data analysis tools, to further elucidate cell signaling pathways during cancer metastasis. Specifically, I will explore other post-translational modifications (besides phosphorylation) during cancer metastasis, design experiments to identify additional key signaling regulators, and probe for substrates phosphorylated by the kinases DBF4, PAK2, and GRK6. Essentially, such research into cancer metastasis could substantially expand our knowledge of intracellular molecules regulations during cancer metastasis, as well as provides information that may lead to the discovery of potential therapeutic targets for preventing cancer metastasis.

References that I co-authored:

1. **Lu, B.**; Chen, T. (2003) A Suboptimal Algorithm for De Novo Peptide Sequencing via Tandem Mass Spectrometry. ***Journal of Computational Biology***. 10: 1-12.

2. **Lu, B.**; Chen, T. (2004) Algorithms for de novo peptide sequencing via tandem mass spectrometry. ***Drug Discovery Today: BioSilico***. 2: 85-90.

3. **Lu, B.**; Chen, T. (2003) A Suffix Tree Approach to the Interpretation of Tandem Mass Spectra: Applications to Peptides of Nonspecific Digestion and Post-translational Modifications. ***Bioinformatics*** Suppl. 2: ii113-ii121.

4. **Lu, B.**; Motoyama, A.; Ruse, C.; Venable, J.; Yates, J.R., III (2008) Improving Protein Identification Sensitivity by Combining MS and MS/MS Information for Shotgun Proteomics Using LTQ-Orbitrap High Mass Accuracy Data. ***Analytical Chemistry***, 80: 2018-2025.

5. **Lu, B.**; Ruse, C.; Xu, T.; Park, S.K.; Yates, J.R. (2007) Automatic Validation of Phosphopeptide Identifications from Tandem Mass Spectra. ***Analytical Chemistry***. 79: 1301-1310.

6. **Lu, B.**; Ruse, C.I.; Yates, J.R., III (2008) Colander: A probability-based support vector machine algorithm for automatic screening for CID spectra of phosphopeptides prior to database search. ***Journal of Proteome Research****.* 7(8): 3628-3634.

7. Ruse, C.I.; McClatchy, D.B.; **Lu, B.**; Cociorva, D.; Motoyama, A.; Park, S.K.; Yates, J.R., III. (**2008**) Motif-specific sampling of phosphoproteomes. ***Journal of Proteome Research.*** 7(5):2140-2150.

8. Cantin, G.T.; Yi, W.; **Lu, B.**; Park, SK; Xu, T.; Lee, JD; Yates, JR, III (2008) Combining protein-based IMAC, peptide-based IMAC, and MudPIT for efficient phosphoproteomic analysis. ***Journal of Proteome Research***. 7: 1346-1351.

9. Chen, Y*.; **Lu, B.***; Yang, Q.; Fearns, C.; Yates, J.R., III; Lee, J.D. (2009) Integrin Phosphoproteomics Identified Key Kinases Regulating Cell Adhesion and Motility. ***Cancer Research****.* 69: 3717-3720.
   * Co-first authors.

10. Liao, L.; McClatchy, D.B.; Park, S.K.; Xu, T.; **Lu, B.**; Yates, J.R., III. (2008) Quantitative analysis of brain nuclear phosphoproteins identifies developmentally regulated phosphorylation events. ***Journal of Proteome Research***. 7: 4743-4755.

11. Chen, R.*; Yang, Q.*; **Lu, B.***; Yi, W.*; Cantin, G.*; Chen, Y.; Fearns, C.; Yates, J.R., III; Lee, J.D. (2009) CDC25B mediates rapamycin-induced oncogenic responses in cancer cells. ***Cancer Research***. 69: 2663-2668.
* Co-first authors.

12. Dimaggio, P.A., Jr; Floudas, C.A.; **Lu, B.**; Yates, J.R., III. (2008) A Hybrid Method for Peptide Identification Using Integer Linear Optimization, Local Database Search, and Quadrupole Time-of-Flight or OrbiTrap Tandem Mass Spectrometry. ***Journal of Proteome Research***, 7: 1584-1593.

13. Ambatipudi, K.; **Lu, B.**; Hagen, F.; Melvin, J.E.; Yates, J.R., III (2009) Quantitative Analysis of Age Specific Variation in the Expression of Human Parotid Salivary Proteins using Multidimensional Protein Identification Technology, In Press (**accepted**), ***Journal of Proteome Research***

14. Koulov, A.*; LaPointe, P.*; **Lu, B.***; Dong, M.Q.;Laister, R.; Matteson, J.; Arrowsmith, C.; Yates, J.R., III; Balch, W.E. (2009) Structural Basis for Aha1 Regulation of Hsp90 ATPase Activity in Cystic Fibrosis (Submitted)
* Co-first authors.

References for which I did not co-author:

Gupta GP, Massague J. Cancer metastasis: building a framework. Cell 2006; 127: 679-95.

Miranti CK, Brugge JS. Sensing the environment: a historical perspective on integrin signal transduction. Nat Cell Biol 2002; 4: E83-90.